Journal of Advanced Research xxx (xxxx) xxx

Contents lists available at ScienceDirect

Journal of Advanced Research

journal homepage: www.elsevier.com/locate/jare

Multi-omics integration identifies *GmISO5* as a key regulator of seed isoflavone biosynthesis and implicates it in the response to soybean mosaic virus

Xinkang Feng [a,b,1], Daqian Sun [a,1], Hui Liu [a], Li Chu [a], JiaJia Li [c], Delin Li [a,d], Hao Zhang [a], Xueqing Wang [a], Tianli Ge [a], Haiyang Zheng [a], Shiyu Guo [a], Xiangguang Zhao [a], Jinyu Li [a], Tuanjie Zhao [b], Yinghui Li [a,*], Xiaobo Wang [c,*], Lijuan Qiu [a,*]

[a] *State Key Laboratory of Crop Gene Resources and Breeding/The National Key Facility for Crop Gene Resources and Genetic Improvement (NFCRI)/ Key Laboratory of Grain Crop Genetic Resources Evaluation and Utilization (MARA)/ Bio-breeding Laboratory of Anhui Province, Institute of Crop Sciences, Chinese Academy of Agricultural Sciences, Beijing 100081, China*

[b] *Soybean Research Institute/National Centre for Soybean Improvement/Key Laboratory for Biology and Genetic Improvement of Soybean (General), Ministry of Agriculture/National Key Laboratory for Crop Genetics and Germplasm Enhancement, Nanjing Agricultural University, Nanjing 210095, China*

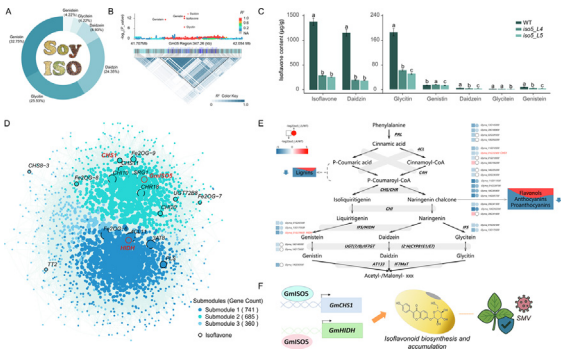
[c] *School of Agronomy, Anhui Agricultural University, Bio-breeding Laboratory of Anhui Province, Hefei 230036, China*

[d] *Xianghu Laboratory, Hangzhou 311231, China*

HIGHLIGHTS

- GWAS/TWAS integration identified GmISO5 as a seed isoflavone locus.
- CRISPR mutants showed an ~80% decrease in total seed isoflavones.
- GmISO5 directly activates *GmHIDH* and *GmCHS1* in the isoflavone pathway.
- Loss of *GmISO5* was accompanied by reduced leaf isoflavone accumulation and increased susceptibility to SMV.

GRAPHICAL ABSTRACT



ARTICLE INFO

Article history:
Received 10 March 2026
Revised 4 September 2026
Accepted 12 September 2026
Available online xxxx

Keywords:
Soybean
Isoflavone biosynthesis
GmISO5
Soybean mosaic virus
Molecular breeding

ABSTRACT

Introduction: Isoflavones are specialized metabolites that accumulate abundantly in soybean seeds, contributing to nutritional quality, human health benefits, and plant defense. Although the core isoflavone biosynthetic pathway has been well characterized, its upstream regulatory network and potential relationship with plant stress responses remain incompletely understood.

Objectives: This study aimed to identify key regulators of seed isoflavone accumulation and to assess whether the identified regulator is associated with the soybean response to soybean mosaic virus (SMV) infection.

Methods: We integrated genome-wide and transcriptome-wide association analyses with CRISPR/Cas9-mediated knockout, transcriptome profiling, molecular interaction assays, co-expression network analysis, and population genetic analysis.

Results: *GmISO5* was identified as a major genetic determinant of seed isoflavone content. CRISPR/Cas9-mediated knockout of *GmISO5* reduced total seed isoflavones by approximately 80% and broadly altered the expression of isoflavone-pathway genes. Molecular assays demonstrated that GmISO5 directly binds

* Corresponding authors.
E-mail addresses: liyinghui@caas.cn (Y. Li), wangxiaobo@ahau.edu.cn (X. Wang), qiulijuan@caas.cn (L. Qiu).
[1] These authors contributed equally to this work.

https://doi.org/10.1016/j.jare.2026.09.011

Please cite this article as: X. Feng, D. Sun, H. Liu et al., Multi-omics integration identifies *GmISO5* as a key regulator of seed isoflavone biosynthesis and implicates it in the response to soybean mosaic virus, Journal of Advanced Research, https://doi.org/10.1016/j.jare.2026.09.011

to and activates the promoters of *GmHIDH* and *GmCHS1*, which encode 2-hydroxyisoflavanone dehydratase and chalcone synthase, respectively, demonstrating direct regulation at distinct steps of the isoflavone biosynthetic pathway. Transcriptome profiling and co-expression network analysis further placed *GmISO5* within an isoflavone-enriched regulatory module. Following SMV inoculation, *GmISO5* knockout lines exhibited more severe symptoms and higher viral accumulation, implicating *GmISO5* in the response to SMV infection. Population genetic analyses showed that the low-isoflavone haplotype ($GmISO5^{Low}$) reached 91% frequency in improved cultivars, suggesting that its enrichment may have resulted from indirect selection for seed-weight and oil-content traits associated with $GmISO5^{Low}$ rather than direct selection for reduced isoflavone accumulation.

Conclusion: GmISO5 is a key regulator of soybean seed isoflavone accumulation that directly activates *GmHIDH* and *GmCHS1*. The increased SMV susceptibility of *GmISO5* knockout lines further implicates *GmISO5* in the soybean response to SMV infection, although whether reduced isoflavone accumulation directly causes the altered SMV phenotype remains to be established.

Introduction

Soybean (*Glycine max* (L.) Merr.) is a premier source of plant oil and protein, serving as a vital industrial raw material and a cornerstone of livestock feed io [1]. Beyond these primary roles, soybean is distinguished by its high accumulation of isoflavones-specialized metabolites that offer significant human health benefits [2]. These compounds comprise twelve structurally distinct derivatives originating from three conserved core structures [3,4]. While daidzein- and genistein-type isoflavones are prized for their antioxidant and anticancer activities [5,6], glycitein-types are often associated with bitterness in soybean products [7]. Biologically, isoflavones facilitate legume-rhizobia symbiosis and bolster defense against pathogens [3,8,9]. Thus, elucidating the genetic regulation of isoflavones biosynthesis is crucial for the rational improvement of soybean nutritional quality and stress resilience. More broadly, the sustainable use of natural bioactive resources requires not only functional discovery but also attention to quality control, safety, standardization, and practical applications in food and pharmaceutical fields [10–14].

The isoflavone biosynthetic pathway, branching from the phenylpropanoid and flavonoid pathways [15], is driven by key enzymes such as chalcone synthase (CHS), chalcone isomerase (CHI), and isoflavone synthase (IFS) [16]. Isoflavone content is a complex quantitative trait controlled by multiple genes and environmental factors [17]. Although various transcription factors have been identified, including positive regulators (e.g., *GmMYB176* [18], *GmMYB133* [19], *GmMYB29* [20,21], $GmZFP7^{22}$) and repressors (e.g., *GmMYB77* [23], $GmMYB100^{24}$), as well as light-responsive factors (*GmCRY1s*, *GmPHOT1s*, *GmSTF1*) [25], the causal regulators explaining natural variation in seed isoflavone content remain largely elusive. Recently, multi-omics approaches integrating GWAS, TWAS, eQTL, and co-expression networks have emerged as powerful tools to provide multidimensional insights into complex traits and directly identify candidate genes underlying their molecular architecture [26–28].

Members of the phosphatidylethanolamine-binding protein (PEBP) family, such as FLOWERING LOCUS T (FT), TERMINAL FLOWER 1 (TFL1) and MOTHER OF FT AND TFL1 (MFT), are well-known in diverse plant species for regulating development, including the control of flowering time, inflorescence meristem identity, seed development and germination [29–33]. MFT plays a prominent role in seed germination and stress responses by modulating ABA and GA signaling [30,34]. In soybean, the MFT homolog *GmMFT* (also known as *GmST05*) has been reported as a major locus controlling seed oil and protein content [35,36]. Despite these insights into primary metabolism and development, the roles of PEBP proteins in secondary metabolism remain poorly understood, and no direct link has yet been established between PEBPs and isoflavone biosynthesis.

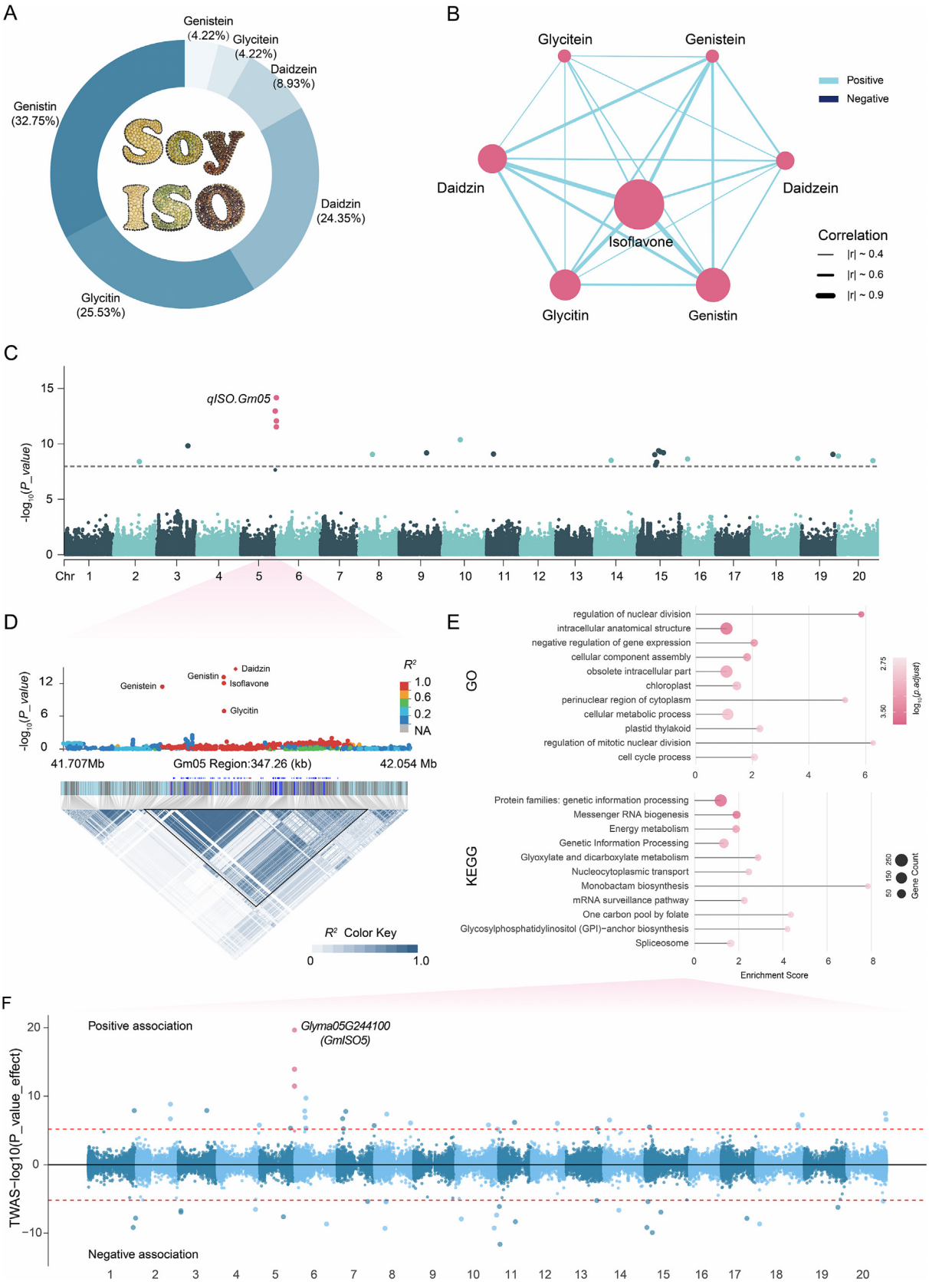
Here, integrative GWAS and TWAS analyses of 4.7 million SNPs identified *GmISO5* as a major locus underlying natural variation in seed isoflavone content. CRISPR/Cas9-mediated disruption of *GmISO5* reduced total seed isoflavone accumulation by approximately 80% and altered the expression of numerous isoflavone-pathway genes. Molecular assays further demonstrated that GmISO5 directly binds to and transactivates the promoters of both *GmHIDH* and *GmCHS1*, revealing direct regulation at multiple positions within the isoflavone pathway. Population genetic analyses indicated that the low-isoflavone haplotype became predominant during soybean domestication and improvement, whereas the high-isoflavone haplotype remains rare. In addition, *GmISO5* knockout lines exhibited increased SMV accumulation and disease severity. These findings establish *GmISO5* as a multi-target regulator of soybean isoflavone biosynthesis and implicate it in the response to SMV infection, although the causal relationship between isoflavone metabolism and SMV susceptibility remains unresolved.

Results

Diversity and heritability of seed isoflavones in soybean

To characterize the natural variation in seed isoflavone composition, we conducted multi-year phenotyping in Beijing (39°N, 116°E) over three consecutive growing seasons (2009–2011) using

Fig. 1. GWAS and TWAS of isoflavone traits. A. Isoflavone content in 311 soybean accessions. Phenotypic data for six components (Genistin, Glycitin, Daidzin, Genistein, Glycitein, and Daidzein) were collected over three years in Beijing, averaged across environments, and analyzed using BLUP. **B.** Correlation network of isoflavone traits. Light blue edges indicate positive correlations, dark blue edges indicate negative correlations, and dot size reflects the relative proportion of each component in total isoflavone content. **C.** GWAS results and QTLs for total isoflavones and six individual components. The x-axis represents chromosomes, and the y-axis shows $-\log_{10}(P)$ values. The dashed line indicates the genome-wide significance threshold. **D.** Local Manhattan plot of the 347.26-kb region surrounding the peak locus *qISO.Gm05*. Colors represent linkage disequilibrium (LD, r^2) between the lead SNP and neighboring variants. **E.** Functional enrichment of isoflavone-associated genes identified under a relaxed threshold ($P < 5 \times 10^{-4}$). GO (up) and KEGG (down) pathway enrichment. The y-axis denotes enriched terms, the count value indicates the number of overlapping genes, and the adjusted P value reflects enrichment significance. **F.** Manhattan plots of TWAS signals for total isoflavones and six individual components (Genistin, Glycitin, Daidzin, Genistein, Glycitein, and Daidzein). Each dot represents a gene; those above or below the horizontal line are positively or negatively associated with isoflavone content, respectively.


A
Genistein (4.22%)
Glycitein (4.22%)
Daidzein (8.93%)
Genistin (32.75%)
Soy ISO
Daidzin (24.35%)
Glycitin (25.53%)
B
Glycitein
Genistein
Positive
Negative
Daidzin
Daidzein
Isoflavone
Correlation
|r| ~ 0.4
|r| ~ 0.6
|r| ~ 0.9
Glycitin
Genistin
C
qISO.Gm05
-log10(P_value)
Chr 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20
D
-log10(P_value)
Genistein
Genistin
Daidzin
Isoflavone
Glycitin
R²
1.0
0.6
0.2
NA
41.707Mb
Gm05 Region:347.26 (kb)
42.054 Mb
R² Color Key
0
1.0
E
GO
regulation of nuclear division
intracellular anatomical structure
negative regulation of gene expression
cellular component assembly
obsolete intracellular part
chloroplast
perinuclear region of cytoplasm
cellular metabolic process
plastid thylakoid
regulation of mitotic nuclear division
cell cycle process
log10(p.adjust)
KEGG
Protein families: genetic information processing
Messenger RNA biogenesis
Energy metabolism
Genetic Information Processing
Glyoxylate and dicarboxylate metabolism
Nucleocytoplasmic transport
Monobactam biosynthesis
mRNA surveillance pathway
One carbon pool by folate
Glycosylphosphatidylinositol (GPI)−anchor biosynthesis
Spliceosome
Gene Count
Enrichment Score
F
TWAS-log10(P_value_effect)
Positive association
Glyma05G244100
(GmISO5)
Negative association
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20

a diverse panel of 311 cultivated soybean accessions [37]. The results showed that glycoside-type isoflavones (genistin, glycitin, and daidzin) accumulated at substantially higher levels than their corresponding aglycones (genistein, glycitein, and daidzein). Among all components, genistin was the most abundant component, accounting for 32.75% of total isoflavones, whereas genistein was the least prevalent (4.23%) (Fig. 1A; Supplementary Table S2). Best linear unbiased prediction (BLUP)-based correlation analysis further revealed that glycoside-type components were the primary contributors to total content and showed the strongest correlations with total isoflavones (Pearson r = 0.71–0.89; Fig. 1B; Supplementary Table S3), while aglycone components showed only moderate correlations. In alignment with these patterns, broad-sense heritability (H^2) estimates indicated that glycoside-type and total isoflavone contents were largely under genetic control, whereas aglycones are more susceptible to environmental factors. Specifically, genistin and total isoflavone content displayed relatively high heritability (H^2 = 0.57 and 0.52, respectively), whereas aglycones exhibited substantially lower heritability (H^2 = 0.12–0.30) (Supplementary Table S4).

From an evolutionary and geographic perspective, population comparisons indicated that landraces contained significantly higher total isoflavones and major individual components than improved cultivars (Fig. S1A), suggesting that modern breeding has unintentionally favored cultivars with reduced isoflavone level. Furthermore, we observed a clear latitudinal line in isoflavone distribution, with higher accumulation in southern accessions compared to northern ones (Fig. S1B).

Considering their low heritability and pronounced environmental sensitivity, aglycone-type isoflavones were treated as secondary traits with moderate confidence in subsequent analyses. Consequently, our primary GWAS focused on glycoside-type isoflavones and total isoflavone content, which provided greater statistical power for detecting robust genetic associations. Because isoflavone profiles are environment-sensitive, the multi-year Beijing phenotypes were analyzed using BLUP values to reduce year-specific noise. Collectively, these findings validate the reliability of our dataset and establish a solid foundation for dissecting the genetic architecture of isoflavone accumulation in soybeans.

Integrative GWAS and TWAS identify key loci underlying seed isoflavone accumulation

We performed a genome-wide association study (GWAS) in this germplasm panel consisting of 4,765,428 SNPs using the Fixed and Random Model Circulating Probability Unification (FarmCPU) [38]. Population structure was controlled using principal components, and genome-wide significance was determined using a Bonferroni-corrected threshold (p < 1.1E-8). To evaluate potential inflation of association signals, we inspected the QQ plots and calculated genomic inflation factors (λGC) for the GWAS results. The λGC values ranged from 0.814 to 0.941 across traits, suggesting that population structure and false-positive inflation were adequately controlled, although the association tests were relatively conservative for some traits. By integrating results for total isoflavones and six individual components, we mapped 22 QTLs (*qISO1-qISO22*) (Fig. 1C, Fig. S2, Supplementary Table S5). Seven overlapped with previously reported QTLs, confirming the robustness of our analysis (Supplementary Table S5). A prominent association was detected at the distal end of chromosome 5, where signals for four isoflavones converged within 74 kb (lead SNPs: *Gm05_41805842*, *Gm05_41867457*, and *Gm05_41879227*; Supplementary Table S5), which overlapped with previously reported QTLs for isoflavone-related traits [39–42]. Linkage disequilibrium analysis placed these SNPs within a single block, narrowing the candidate region to 200.5 kb (Fig. 1D). The interval spans 27 protein-coding genes in the Wm82.a4.v1 reference genome, but pinpointing the causal gene remains challenging. Therefore, we combined these four peak SNPs and designated the locus on chromosome 5 as *qISO.Gm05*.

To link gene expression with isoflavone variation, we conducted a transcriptome-wide association study (TWAS) using seed transcriptomes at the R5.5 stage. Using the exon proportion-based association approach reported by Li et al. [28], we applied the FarmCPU model [38] to analyze 154,119 exons and identified 71 genes significantly associated with isoflavone content after Bonferroni correction ($p < 6.5 \times 10^{-6}$) (Fig. 1F, Fig. S3). Gene Ontology analysis revealed that these genes were enriched in processes related to nuclear division, gene expression regulation, and metabolism, while KEGG analysis highlighted pathways of genetic information processing, energy metabolism, and amino acid metabolism (Fig. 1E). Among the trait-associated genes, *Glyma.05G244100* was repeatedly detected for four isoflavones. This gene comprises four exons, with exon 2 proportion showing a strong positive correlation with seed isoflavones. Importantly, it resides within the major-effect QTL *qISO.Gm05* and was subsequently validated as a candidate gene, hereafter designated *GmISO5*, which was previously reported as *GmST05* [36] and *GmMFT* [35].

Genomic variation of GmISO5 is correlated with seed isoflavone content

We analyzed sequence variation of *GmISO5* in 1566 previously re-sequenced soybean accessions [37] and identified nine variants located in the promoter and exon regions. Linkage disequilibrium grouped these variants into two haplotypes (Fig. 2A; Supplementary Table S2). TWAS of 204 accessions showed that carriers of *GmISO5Hap* [1] exhibited a significantly higher proportion of exon 2 usage than those with *GmISO5Hap* [2] (Fig. 2J), consistent with a sharp reduction in transcript abundance at the end of exon 4 and in the 3′ UTR of Hap2 accessions (Fig. 2B; Supplementary Table S2). Phenotypically, Hap1 accessions accumulated significantly higher levels of total isoflavones and their major constituents, whereas *GmISO5Hap* [2] accessions displayed the opposite trend (Fig. 2C–I; Supplementary Table S2). The altered terminal transcript abundance associated with exon 4 variations (+22 bp and −1 bp) may contribute to the reduced isoflavone content in *GmISO5Hap* [2]. Together, these results establish that *GmISO5Hap* [1] represents the high-isoflavone haplotype (renamed $GmISO5^{High}$), while *GmISO5Hap* [2] corresponds to the low-isoflavone haplotype (renamed $GmISO5^{Low}$).

Functional characterization of GmISO5 in soybean isoflavone biosynthesis

To determine the role of *GmISO5* in isoflavone biosynthesis, we first examined its expression patterns by qRT-PCR. *GmISO5* was specifically expressed in developing seeds, peaking at 30 days after flowering (Fig. S4A). Within seed compartments, expression was concentrated in the cotyledon and embryonic axis but negligible in the seed coat (Fig. S4B). Consistent with these data, the Seed Development Gene Network database (https://seedgenenetwork.net/soybean) indicated high expression in the aleurone, abaxial epidermis, and parenchyma of the cotyledon, as well as in the parenchyma and plumule of the embryonic axis (Fig. S4C), suggesting a regulatory role in seed development and metabolism.

To validate its function, we generated two independent CRISPR/Cas9 knockout lines (*iso5_L4* and *iso5_L5*) in the Wm82 background, which carries $GmISO5^{Low}$ (Hap2). Both carried frameshift deletions (4 bp and 5 bp) in exon 1, producing premature stop codons (Supplementary Table S7). Knockout of *GmISO5* caused dra-

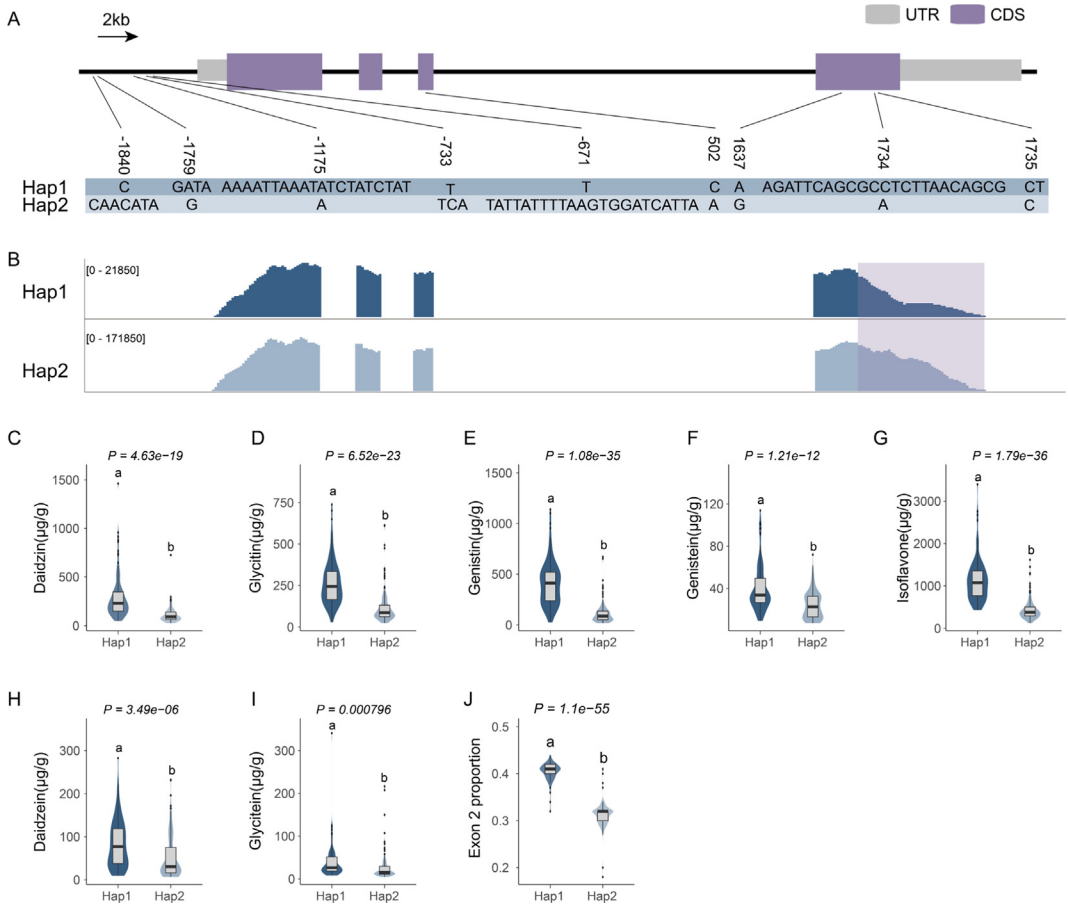


Fig. 2. Allelic effects of *GmISO5* on seed isoflavone content in soybean. A. Gene structure of *GmISO5* and its haplotypes. Major variation sites in the promoter, exonic, and UTR regions are indicated; arrows denote transcriptional direction. **B.** Transcript abundance of *GmISO5* in accessions carrying different haplotypes. **C–I.** Distribution of isoflavone components in accessions with distinct haplotypes: (C) Daidzin, (D) Glycitin, (E) Genistin, (F) Genistein, (G) total isoflavones, (H) Daidzein, and (I) Glycitein (μg/g). **J.** Proportion of exon 2 in accessions with different haplotypes.

matic reductions in glycitin, daidzin, daidzein, glycitein, and genistein, with only a marginal increase in genistin. Total seed isoflavones dropped from 1375.52 μg/g in the wild type to 273.67 μg/g in the mutants, representing an 80.1% reduction (Fig. 3A; Supplementary Table S8), establishing GmISO5 as a critical regulator of isoflavone accumulation.

To gain mechanistic insight, we performed RNA-seq on R5.5 stage seeds from Wm82 and *iso5* mutants, with three biological replicates per genotype. Over 85% of reads mapped uniquely to the reference genome (Supplementary Table S9), and Pearson correlation coefficients among biological replicates were as high as 0.967, confirming the reproducibility of the RNA-seq data (Fig. S5A). Principal component analysis revealed clear separation among WT, *iso5_L4*, and *iso5_L5* (Fig. S5B). In total, 4902 and 2682 genes were differentially expressed ($|\log_2\text{fold change}| \geq 1$, FDR < 0.05) in *iso5_L4* and *iso5_L5*, respectively (Fig. 3B; Fig. S6; Supplementary Table S10). GO enrichment showed strong overrepresentation of isoflavone biosynthetic and metabolic processes (Fig. 3C). Strikingly, 30 of 65 known isoflavone biosynthetic genes were regulated by *GmISO5*, including multiple enzymes catalyzing key steps in the pathway (Fig. 3D, E; Supplementary Table S10). Expression changes of eight representative genes were further confirmed by qRT-PCR, consistent with RNA-seq results (Fig. S7). Together, these findings demonstrate that *GmISO5* is seed-specific, essential for isoflavone accumulation, and exerts broad transcriptional control over the biosynthetic network.

GmISO5 directly activates GmHIDH and GmCHS1 transcription

GmCHS1 encodes a chalcone synthase that catalyses an early step in flavonoid and isoflavonoid biosynthesis by generating chalcone precursors and has also been identified as a key component of the soybean isoflavonoid metabolon [43]. *GmHIDH* encodes a 2-hydroxyisoflavanone dehydratase that catalyses the dehydration of 2-hydroxyisoflavanones to produce the isoflavone aglycones daidzein and genistein [16]. Transcriptome profiling indicated that disruption of GmISO5 affected multiple genes across the isoflavone biosynthetic pathway, among which *GmCHS1* and *GmHIDH* showed pronounced reductions in transcript abundance in both *iso5_L4* and *iso5_L5*. qRT-PCR analysis further confirmed that the expression levels of *GmCHS1* and *GmHIDH* were significantly lower in the two knockout lines than in the wild type (Fig. 4A, B), supporting the positive regulation of both genes by GmISO5.

To determine whether GmISO5 associates with the *GmHIDH* and *GmCHS1* promoters, we performed yeast one-hybrid assays using their promoter fragments as baits. All transformed yeast combinations grew normally on SD-TL medium. On SD-TLH medium supplemented with 50 mM 3-AT, yeast co-transformed with pGADT7-GmISO5 and either the *proGmHIDH* or *proGmCHS1* bait construct retained growth, whereas yeast carrying the corresponding promoter bait and the empty pGADT7 vector failed to grow (Fig. 4C). These results indicate that GmISO5 is associated with both promoters in yeast.

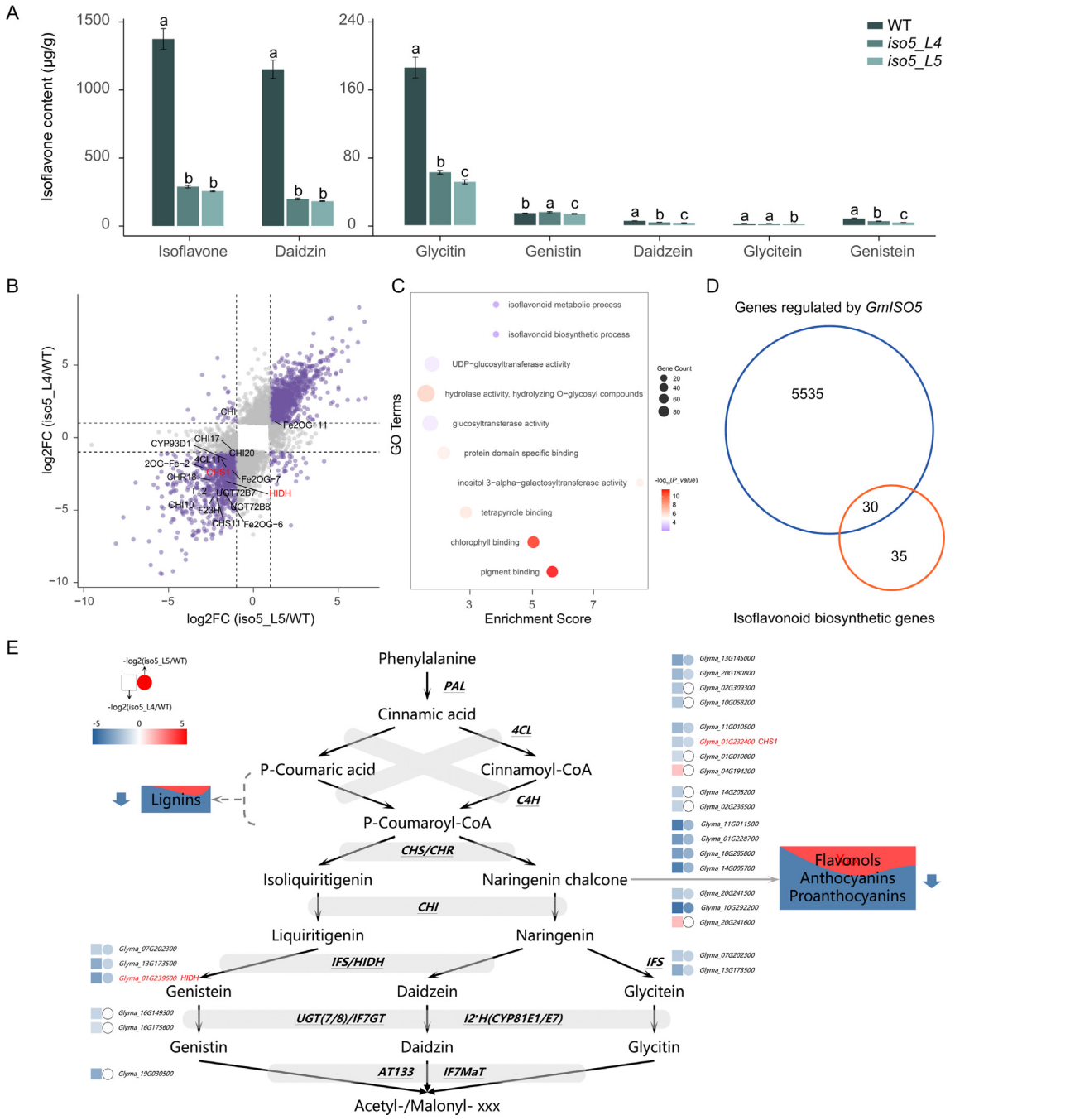


Fig. 3. Regulatory role of *GmISO5* in seed isoflavone accumulation and biosynthesis. A. Isoflavone and component levels in mature seeds of WT, *iso5_L4*, and *iso5_L5*. **B.** Scatterplot of differentially expressed genes (DEGs) in *iso5_L4* vs. WT (y-axis) and *iso5_L5* vs. WT (x-axis). Each point represents a DEG ($log_2FC > 1$). The lines mark known isoflavone biosynthetic genes. **C.** GO enrichment of *GmISO5*-regulated genes. Dot color reflects enrichment significance, and dot size indicates the number of genes. **D.** Venn diagram showing the overlap between *GmISO5*-regulated genes and isoflavone biosynthetic pathway genes. **E.** Schematic representation of the soybean isoflavone biosynthetic pathway. Thirty *GmISO5*-regulated genes identified from *iso5_L4* vs. WT and *iso5_L5* vs. WT comparisons are listed on the left and right, respectively.

We next examined the transcriptional activity of GmISO5 using dual-luciferase reporter assays in *Nicotiana benthamiana* leaves. The *GmHIDH* and *GmCHS1* promoter fragments were independently fused to the LUC reporter, with GmISO5 expressed as the effector (Fig. 4D). Co-expression of GmISO5 with either *proGmHIDH* or *proGmCHS1* produced markedly stronger luminescence signals and significantly higher LUC/REN ratios than the corresponding control combinations (Fig. 4E, F), indicating that GmISO5 transactivates both promoters.

We subsequently performed electrophoretic mobility shift assays to examine direct and sequence-specific DNA binding. Recombinant GmISO5 bound to two *GmHIDH* promoter fragments located at −732 to −678 bp and −641 to −583 bp relative to the start codon (Fig. 4G, I, J), as well as to a *GmCHS1* promoter fragment located at −845 to −789 bp (Fig. 4H, K). In each case, the shifted protein-DNA complexes were progressively weakened by 10- and 50-fold molar excesses of the corresponding unlabeled probes, whereas binding was substantially reduced or abolished when mutated probes were used. Sequence comparison of the three EMSA-positive fragments revealed a shared A/T-rich sequence arrangement, TTAAAA-N4-6-ATAAT, in which the two conserved elements occurred in the same orientation and were separated

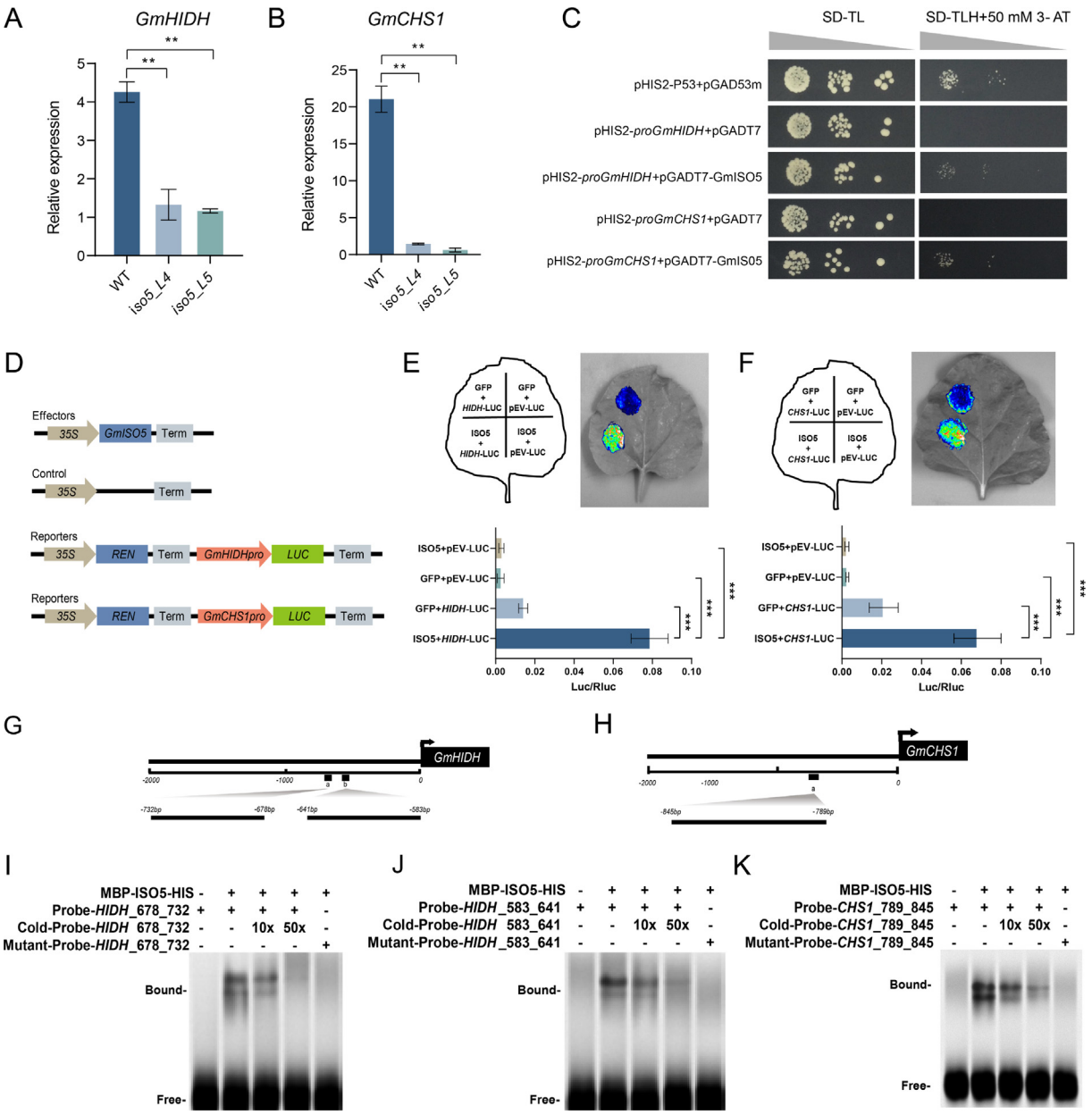


Fig. 4. GmISO5 directly binds to and activates the promoters of *GmHIDH* and *GmCHS1*. A–B. qRT-PCR analysis of *GmHIDH* (A) and *GmCHS1* (B) expression in developing seeds of wild type, *iso5_L4*, and *iso5_L5*. **C**. Yeast one-hybrid assays showing the association of GmISO5 with the *GmHIDH* and *GmCHS1* promoters. **D**. Schematic representation of the effector and reporter constructs used in the dual-luciferase assays. E-F. Dual-luciferase assays showing the activation of the *GmHIDH* (E) and *GmCHS1* (F) promoters by GmISO5 in *Nicotiana benthamiana* leaves. Representative luminescence images and quantitative LUC/REN ratios are shown. **G–H**. Schematic diagrams of the *GmHIDH* (G) and *GmCHS1* (H) promoters showing the positions of the EMSA probes relative to the start codons. **I–J**. EMSA showing direct and sequence-specific binding of recombinant GmISO5 to two regions of the *GmHIDH* promoter, spanning −732 to −678 bp (I) and −641 to −583 bp (J). **K**. EMSA showing direct and sequence-specific binding of recombinant GmISO5 to the −845 to −789 bp region of the *GmCHS1* promoter. Unlabeled wild-type probes were added at 10-fold or 50-fold molar excesses as competitors. Mutated probes were used to assess binding specificity. Bound protein-DNA complexes and free probes are indicated.

by four to six nucleotides. This sequence arrangement may contribute to promoter recognition by GmISO5, although additional analyses are required to determine whether it represents a general GmISO5-binding motif.

Collectively, the qRT-PCR, yeast one-hybrid, dual-luciferase, and EMSA results identify *GmHIDH* and *GmCHS1* as direct transcriptional targets of GmISO5. As these genes function at distinct positions in the isoflavone biosynthetic pathway, their combined regulation provides a plausible molecular explanation for the broad reduction in isoflavone accumulation following *GmISO5* disruption. Other pathway genes affected in the RNA-seq analysis may represent additional direct or indirect downstream components of the GmISO5 regulatory network and require further experimental validation.

Selection patterns of GmISO5 haplotypes and gene modules coordinating isoflavone accumulation in soybean

Given the nutritional and breeding importance of isoflavone content in soybeans, we next investigated whether natural variation at *GmISO5* was subject to artificial selection during domestication and improvement. Analysis of 1566 re-sequenced accessions (56 wild soybeans and 1510 cultivars) revealed two major haplotypes: a high-isoflavone haplotype (Hap1, $GmISO5^{High}$) and a low-

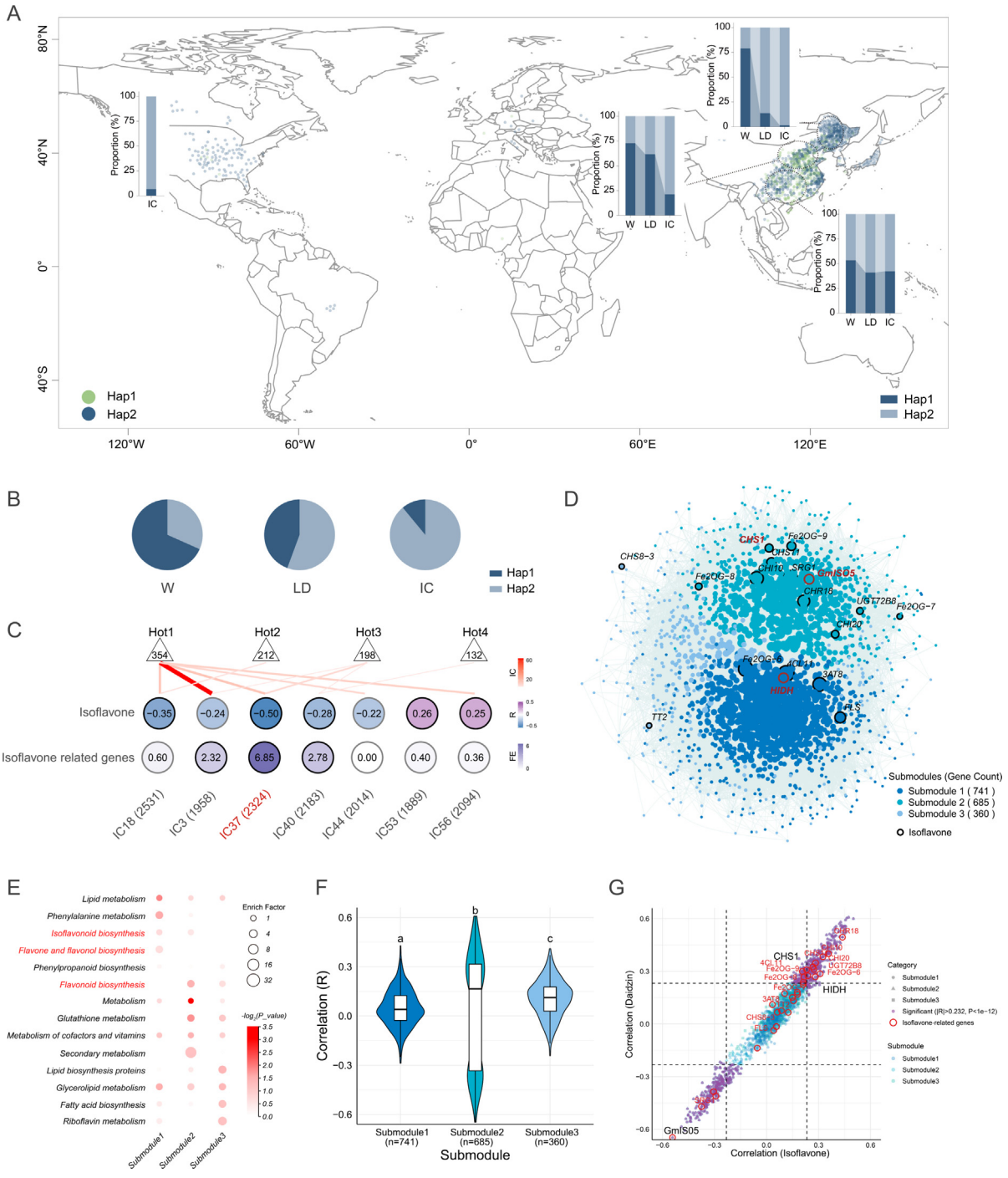


Fig. 5. Selection patterns and regulatory networks associated with *GmISO5* in soybean. A. Global distribution of *GmISO5* haplotypes. Each dot represents an accession, with colors indicating haplotype identity. Bar plots depict haplotype frequencies in major production regions (USA and China). **B.** Frequency distribution of the two haplotypes—Hap1 (high-isoflavone) and Hap2 (low-isoflavone)—in *G. soja* (W), landraces (LD), and cultivars (IC). **C.** Gene co-expression modules significantly associated with isoflavone content. The first heatmap row shows module-trait correlations, and the second row shows fold enrichment of known isoflavone-related genes. The width and intensity of red lines indicate the strength of distal eQTL hotspot-mediated regulation, with the number of regulated genes labeled on triangles. **D.** Gene network of the IC37 module, highlighting isoflavone biosynthesis and metabolism genes. **E.** Overrepresented KEGG pathways in the three IC37 submodules. **F.** Correlation between gene expression and isoflavone content across IC37 submodules. Boxplots show median (center line) and interquartile range (box boundaries). **G.** Scatter plot of gene-wise correlations with total isoflavone content (x-axis) and daidzin content (y-axis). Genes significantly correlated with both traits ($P < 1 \times 10^{-12}$, Pearson) are highlighted in purple.

isoflavone haplotype (Hap2, $GmISO5^{Low}$) (Supplementary Table S2) [37]. The frequency of $GmISO5^{Low}$ increased sharply from *G. soja* to landraces and reached 91% in improved cultivars (Fig. 5B), indicating that $GmISO5^{Low}$ underwent strong positive selection, whereas $GmISO5^{High}$ was progressively eliminated during domestication and breeding.

To further evaluate whether *GmISO5* was subject to artificial selection during soybean domestication and improvement, we examined population genetic signals across a 400-kb genomic region encompassing the *GmISO5* locus (41.67–42.07 Mb) using resequencing data. Nucleotide diversity (π) analysis revealed that wild soybeans exhibited substantially higher genetic diversity in

this region than landraces, whereas improved cultivars showed a pronounced reduction in π values (Fig. S8), consistent with a strong selective sweep. Tajima's *D* values were positive and relatively high in landraces, suggesting the presence of balancing selection or the maintenance of multiple haplotypes at an intermediate stage of domestication. In contrast, Tajima's *D* values were markedly reduced in improved cultivars, indicative of recent strong positive selection contributing to the pronounced enrichment of $GmISO5^{Low}$. Consistently, F_{st} analysis revealed elevated genetic differentiation between wild, landrace, and cultivated populations across this region (Fig. S8), further supporting the action of selection at the *GmISO5* locus. Together, these population genetic signatures indicate that *GmISO5* experienced at least two major evolutionary transitions during soybean domestication and improvement, ultimately leading to the preferential enrichment of the low-isoflavone haplotype in improved cultivars.

Geographical distribution analysis further demonstrated that $GmISO5^{Low}$ is highly enriched in improved cultivars from major soybean production regions, particularly the United States and northern China, with a clear latitudinal increase in frequency across China (Fig. 5A). These findings highlight the underutilization of the high-isoflavone $GmISO5^{High}$ haplotype in contemporary breeding programs, particularly in northern China and the US, underscoring its potential value for nutritional improvement.

To gain mechanistic insight into the regulatory architecture of seed isoflavone variation, we constructed a gene co-expression network using independent component analysis (ICA). Among 58 modules identified, seven were significantly correlated with isoflavone content (Fig. 5C; Supplementary Table S11). We selected the four eQTL hotspots with the largest numbers of distally regulated genes because these loci had the greatest potential to coordinate trans-regulatory variation. Other eQTL hotspots regulated fewer genes and showed weaker relationships with isoflavone-associated modules (Supplementary Table S12). Notably, module IC37 showed the strongest association with isoflavone content. Its expression-correlation network revealed an interconnected structure formed by genes involved in isoflavone biosynthesis and metabolism, including $CHS1^{44}$, $CHS11^{44}$, CHR^{45}, *CHI*, $4CL^{46}$, *UGT72B8*, *Fe2OG*, and $TT2^{47}$ (Fig. 5D, G). Importantly, this module also contained *GmISO5* and its experimentally validated regulatory target, *GmHIDH* (Fig. 5C, D), supporting the reliability of module-based network mining and suggesting that additional genes within IC37 may participate in the *GmISO5*-associated isoflavone regulatory network. To further resolve the functional organization of IC37, we divided it into three submodules with distinct enrichment patterns (Fig. 5E; Supplementary Table S13). Submodule 1 was enriched in lipid metabolism and flavonoid-related processes, Submodule 3 was mainly associated with lipid and fatty acid biosynthesis, whereas Submodule 2 was specifically enriched for flavonoid/isoflavonoid biosynthesis and secondary metabolism. Consistently, Submodule 2 showed the strongest correlation with seed isoflavone content (Fig. 5F), indicating that it may represent the core functional subnetwork within IC37 that contributes most directly to isoflavone accumulation.

Collectively, these results show that *GmISO5* haplotypes experienced distinct selection patterns during soybean domestication and improvement, and that *GmISO5* is embedded within the isoflavone-associated IC37 module. Submodule analysis further identified Submodule 2 as the core isoflavone-associated subnetwork. Because co-expression analysis is correlative, IC37 and its submodules should be interpreted as sources of prioritized candidate genes and potential regulatory associations for future validation.

GmISO5 knockout lines exhibit increased susceptibility to SMV

Isoflavones have been implicated in plant-pathogen interactions see [17,48–50]), and previous studies have also suggested possible antiviral activity of isoflavonoid-related compounds [51,52]. We therefore examined whether disruption of *GmISO5* was associated with an altered response to SMV infection. In Wm82, qRT-PCR showed that the transcript levels of *GmISO5* and *GmHIDH* increased at 2 dpi after SMV inoculation (Fig. 6B, C). Following inoculation, *iso5_L4* and *iso5_L5* exhibited more severe symptoms than wild type (Fig. 6D, F, G and showed significantly higher viral accumulation as measured by DAS-ELISA ($P < 0.01$; Fig. 6E). Metabolic profiling further showed that total isoflavone content and several major isoflavone components were reduced in mutant leaves compared with wild type (Fig. 6A). Together, these results show that loss of *GmISO5* was accompanied by reduced leaf isoflavone accumulation and increased susceptibility to SMV. However, these experiments do not establish that the increased viral susceptibility was directly caused by reduced isoflavone accumulation.

Discussion

GmISO5 directly regulates multiple nodes of the isoflavone biosynthetic pathway

Isoflavone accumulation in soybean is controlled by a complex biosynthetic and regulatory network, but functional validation of seed-specific upstream regulators remains limited [23]. Previous studies have primarily focused on canonical transcription factors, including several MYB family members [19,20,24,53,54] and the C2H2-type zinc-finger protein GmZFP7 [22]. In contrast, GmISO5 belongs to the PEBP family, whose characterized functions have mainly involved flowering, seed development, germination, and stress responses [29–34]. Therefore, the identification of GmISO5 expands the known regulatory repertoire of soybean isoflavone biosynthesis beyond previously characterized MYB and zinc-finger regulators.

Mechanistically, we identified *GmHIDH* and *GmCHS1* as two direct transcriptional targets of GmISO5. *GmCHS1* encodes chalcone synthase, which acts at an early step of flavonoid biosynthesis, whereas *GmHIDH* encodes 2-hydroxyisoflavanone dehydratase, which functions in the isoflavone-specific branch [16,55–57]. Direct regulation of genes at these distinct pathway positions suggests that GmISO5 promotes isoflavone accumulation through multiple regulatory nodes rather than through *GmHIDH* alone. This multi-target regulation provides a plausible explanation for the broad reduction in isoflavone components and the extensive transcriptional changes observed in *GmISO5* knockout seeds. Knockout of *GmISO5* caused a markedly stronger reduction in daidzin-type compounds than in genistin- and glycitin-type compounds. This uneven compositional response may reflect the combined regulation of precursor supply through *GmCHS1* and reactions within the isoflavone-specific branch through *GmHIDH*, although direct evidence from metabolic-flux analysis is still lacking. Because daidzein- and genistein-derived isoflavones have been associated with antioxidant and anticancer properties [5,6], whereas glycitein-related compounds may contribute to bitterness [7], *GmISO5*-mediated changes in isoflavone composition may also influence the nutritional and sensory properties of soybean seeds. Nevertheless, although 30 of the 65 annotated isoflavone-pathway genes were differentially expressed, only *GmHIDH* and *GmCHS1* were directly validated by qRT-PCR, Y1H, Dual-LUC, and EMSA. The remaining genes may represent direct targets or indirect

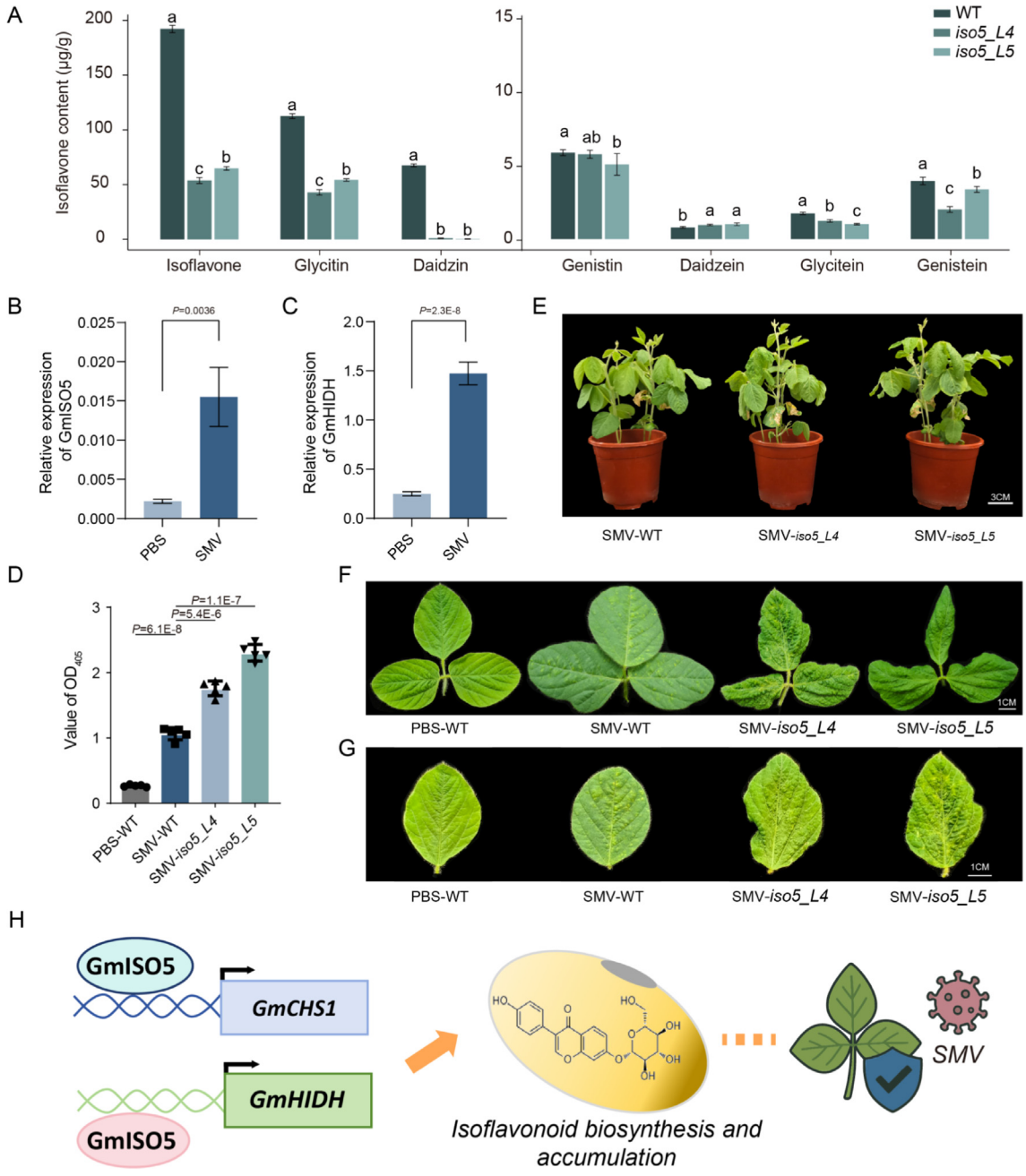


Fig. 6. ***GmISO5*** **knockout is associated with reduced leaf isoflavone accumulation and increased susceptibility to SMV. A.** Isoflavone and component levels in the second trifoliolate leaves of wild type (WT), *iso5_L4*, and *iso5_L5*. **B–C.** qRT-PCR analysis of *GmISO5* (B) and *GmHIDH* (C) expression in Wm82 after inoculation with SMV or PBS (control). **D.** DAS-ELISA quantification of SMV accumulation in wild-type (WT, Wm82) and *iso5_L4*/*iso5_L5* mutants at 2 weeks post inoculation. **E–G.** Phenotypic comparison of WT and mutants following SMV infection: whole plants (E), trifoliate leaves (F), and individual leaves (G). **H.** Proposed model showing that GmISO5 directly activates *GmCHS1* and *GmHIDH* to promote isoflavonoid biosynthesis and accumulation. The dashed line indicates a proposed association between isoflavonoid accumulation and the soybean response to SMV infection, for which the causal mechanism remains unresolved.

downstream responses and require further experimental validation.

Integrating multi-omics to dissect complex regulatory networks

Integrative multi-omics approaches have become powerful tools for dissecting complex crop traits because they connect population-level genetic variation with transcriptomic, epigenomic, metabolic, and network-level evidence [28,58,59]. In crops such as *Brassica napus*, soybean, and cotton, the integration of GWAS, TWAS, population transcriptomes, methylomes, and functional validation has facilitated the identification of candidate genes controlling seed oil content [59], fatty acid composition, glucosinolate accumulation [60], seed coat traits [61], seed weight [62], oil-to-protein ratio [1], and fiber-related traits [63]. These studies demonstrate that multi-omics frameworks can strengthen candidate-gene discovery by linking genetic associations with expression regulation, network context, and functional evidence [59,61,63].

In this study, we extended this integrative strategy to soybean isoflavone metabolism. GWAS provided population-scale DNA-level association evidence, but its resolution may be limited by

linkage disequilibrium and therefore often identifies candidate genomic regions rather than a single causal gene. By contrast, TWAS linked transcript abundance and exon-usage variation to seed isoflavone phenotypes, providing an expression-based layer for candidate-gene prioritization [28]. The convergence of GWAS and TWAS signals supported *GmISO5* as the only co-supported candidate gene within *qISO.Gm05* (Fig. 1D, F).

Network analysis further strengthened this evidence chain. Both *GmISO5* and its experimentally validated downstream target *GmHIDH* were assigned to the IC37 module, together with core isoflavone biosynthetic pathway genes such as *CHS* [44], *CHR* [45], *4CL* [46], and *TT2* [47] (Fig. 5D). This co-expression pattern suggests that *GmISO5*-associated isoflavone regulation is embedded within a broader pathway-level transcriptional network. Submodule analysis provided additional resolution, as Submodule 2 showed the strongest correlation with seed isoflavone content and was specifically enriched for flavonoid/isoflavonoid biosynthesis and secondary metabolism. Together, these results support IC37, particularly Submodule 2, as a useful network framework for prioritizing additional genes involved in *GmISO5*-associated isoflavone regulation.

Soybean isoflavone accumulation is influenced by both developmental stage and environmental conditions. In this study, the TWAS was based on transcriptomic data collected from seeds at the R5.5 developmental stage under a single field environment, whereas the GWAS used multi-year phenotypic BLUP values to reduce year-specific environmental variation. The convergence of GWAS and TWAS at the same major locus therefore provided complementary evidence supporting the prioritization of *GmISO5* as a regulator of isoflavone accumulation during R5.5 seed development. However, the current TWAS may not capture genes acting at other stages of seed development or regulatory loci whose effects are restricted to specific environmental conditions. Future studies integrating matched phenotypic, transcriptomic, and metabolomic data across multiple developmental stages and environments will be required to identify additional isoflavone-related loci influenced by genotype-by-environment interactions.

Evolutionary selection of GmISO5 haplotypes and its association with SMV response

Natural variation at *GmISO5* showed pronounced signatures of domestication, improvement, and geographic differentiation. The frequency of the low-isoflavone haplotype $GmISO5^{Low}$ increased progressively from wild soybean to landraces and improved cultivars, reaching 91% in modern cultivars, whereas the high-isoflavone haplotype $GmISO5^{High}$ became increasingly rare (Fig. 5A, B). $GmISO5^{Low}$ was particularly enriched in major soybean-producing regions, including Northeast China and the United States.

This marked shift in haplotype frequency may reflect historical selection for seed oil content and 100-seed weight rather than direct selection for reduced isoflavone accumulation. Previous studies identified *GmISO5*, also known as *GmMFT* [35] or *GmST05* [36], as a locus affecting seed oil content, seed size, and 100-seed weight. Consistent with these reports, accessions carrying $GmISO5^{Low}$ in the present population had significantly higher 100-seed weight and seed oil content than those carrying $GmISO5^{High}$ (Fig. S9A–C), whereas $GmISO5^{Low}$ was associated with lower seed isoflavone accumulation (Fig. 2C–I). Because seed size and oil content have long been major agronomic and quality targets in soybean breeding [35,36,62], historical selection for large-seeded and high-oil accessions may have indirectly increased the frequency of $GmISO5^{Low}$, with reduced isoflavone accumulation retained as a correlated feature of this haplotype. Under this interpretation, low isoflavone content was not necessarily a direct breeding target, and the decline of $GmISO5^{High}$ may instead have occurred as a correlated consequence of selection for favorable seed traits associated with $GmISO5^{Low}$. Nevertheless, demographic history, regional adaptation, linked variation, and genetic-background effects may also have contributed to the observed haplotype distribution [37].

Isoflavones and related pathway genes have been implicated in diverse biotic interactions in soybean [21,48–50]. More specifically, multiplex CRISPR/Cas9-mediated editing of flavonoid-pathway genes increased soybean isoflavone accumulation and was accompanied by enhanced resistance to SMV [51]. In another plant-virus system, an isoflavonoid isolated from *Nicotiana tabacum* exhibited anti-TMV activity [52]. In the present study, GmISO5 directly activated *GmCHS1* and *GmHIDH*, and loss of GmISO5 reduced isoflavone accumulation in both seeds and leaves. Following SMV inoculation, *GmISO5* knockout lines exhibited more severe symptoms and significantly higher viral accumulation than wild type (Fig. 6). These results implicate *GmISO5* in the soybean response to SMV infection and are consistent with a possible contribution of isoflavone metabolism to antiviral responses. However, they do not establish that reduced isoflavone accumulation directly caused the increased SMV susceptibility of the knockout lines, as other *GmISO5*-regulated metabolic or defense pathways may also contribute to the phenotype. Genetic complementation, *GmISO5* overexpression, metabolite-rescue experiments, and SMV evaluation of contrasting natural haplotypes will be required to clarify the causal relationship.

Conclusion

In conclusion, this study identifies GmISO5 as a PEBP-family regulator underlying natural variation in soybean seed isoflavone accumulation. By integrating GWAS, TWAS, transcriptome and co-expression analyses, CRISPR/Cas9 validation, metabolite profiling, and molecular assays, we demonstrate that GmISO5 directly binds to and activates the promoters of *GmHIDH* and *GmCHS1*, thereby regulating distinct positions of the isoflavone biosynthetic pathway. Haplotype and selection analyses suggested that the enrichment of $GmISO5^{Low}$ may have resulted from indirect selection for its associated high seed weight and oil content, leading to the correlated depletion of the high-isoflavone $GmISO5^{High}$ haplotype. *GmISO5* knockout lines also exhibited more severe SMV symptoms and increased viral accumulation, implicating *GmISO5* in the soybean response to SMV infection. However, whether reduced isoflavone accumulation directly causes the altered SMV phenotype remains unresolved, and genetic complementation, overexpression, and metabolite-rescue experiments will be required to establish the underlying mechanism. Collectively, these findings expand the regulatory framework of soybean isoflavone biosynthesis and provide a foundation for the targeted utilization of *GmISO5* alleles in seed-quality improvement.

Materials and methods

Plant materials and growth conditions

A total of 311 soybean accessions were selected from a previously re-sequenced germplasm collection [37], and detailed accession information is provided in Supplementary Table S1. All accessions were cultivated in Beijing, China (39°N, 116°E) for three consecutive years (2009, 2010, and 2011), during which mature seeds were harvested for isoflavone profiling. Each accession was planted in two replicated rows with 0.10 m spacing between plants and 0.10 m between rows under standard field management practices, including basal fertilization and pest and weed control when

needed according to local agronomic recommendations. Best linear unbiased prediction (BLUP) values for total and component isoflavone contents were calculated using mixed linear models and subsequently used for GWAS.

For transient assays, *Nicotiana benthamiana* plants were grown in a greenhouse at 25 °C under long-day conditions (14 h light/10 h dark, light intensity 524μmol m^{-2} s^{-1}). For functional characterization in soybean, *GmISO5* knockout lines (*iso5_L4* and *iso5_L5*) were sown each June at the Shunyi Experimental Station of the Institute of Crop Sciences, Chinese Academy of Agricultural Sciences (Beijing, 39°N, 116°E). Plants were grown in rows 3 m in length with 0.10 m spacing between rows and 0.10 m between plants, following a randomized block design with standard field management. Developing seeds were sampled at the indicated reproductive stages, and mature seeds were collected after full physiological maturity.

Isoflavone extraction and determination

Isoflavones were extracted from soybean seeds and leaves following the method of Sun et al., with minor modifications [64]. Samples were ground into fine powder with a laboratory cyclone mill (A10 basic; IKA, Staufen, Germany). For each assay, 0.1 g of powder was suspended in 20 mL of 90% (v/v) methanol–water, vortexed, and sonicated in a 60 °C water bath for 30 min. After centrifugation at 12,000 g for 5 min, the supernatant was collected, and the extraction was repeated once. The combined extracts were concentrated under nitrogen to 20 mL, diluted to 25 mL with methanol, centrifuged, and filtered through a 0.22 μm organic-phase membrane filter. Filtrates were stored at –20 °C until analysis.

Isoflavone quantification was performed on an Agilent HPLC system using an Agilent Poroshell 120 EC-C18 column (4.6 × 250 m m, 4 μm) at 25 °C. The mobile phase consisted of solvent A (0.1% trifluoroacetic acid in water) and solvent B (acetonitrile) with a flow rate of 0.8 mL/min and 10 μL injection volume. The gradient was programmed as follows: 0–4 min, 90% A; 4–10 min, linear gradient to 75% A; 10–12 min, 70% A; 12–25 min, 45% A; 25–26 min, back to 90% A; 26–29 min, 90% A. Detection was performed with a diode array detector at 254 nm. Standard curves were prepared using six isoflavones (daidzin, genistin, glycitin, daidzein, genistein, and glycitein; Zhenzhun Biotech, China) at concentrations of 0.25–50 μg/mL. All samples were analyzed in triplicate, and the total isoflavone (TIF) content was calculated as the sum of the six components.

Genotype data processing and association analysis

Previously published resequencing data were used for genotype calling [37]. SNPs were imputed using Beagle and filtered with the following criteria: minor allele frequency (MAF) > 5%, missing rate < 20%, and heterozygosity < 20%. A total of 4,765,428 high-quality SNPs were retained for downstream analyses. GWAS was performed using the FarmCPU model implemented in GAPIT for R [38]. Population structure was controlled by including the top three principal components (PCs) estimated with PLINK v1.90 [65]. Genome-wide significance thresholds were determined by Bonferroni correction at $\alpha = 0.05$ ($p < 1.1E\text{-}8$). QQ plots and genomic inflation factors (λGC) were used to evaluate potential systematic inflation of association signals and false-positive control. QTLs were defined by grouping significant SNPs within the local LD interval or within adjacent physical windows supporting the same trait signal; overlapping QTLs across traits were merged into a single locus.

Transcriptome data and TWAS analysis

RNA-seq data for 204 accessions were obtained from a previously published population-level seed transcriptome dataset [27]. Fresh seeds were collected at the R5.5 stage in 2021 in Sanya, Hainan (18.2°N, 109.9°E). Total RNA extraction, library construction, and sequencing were performed by Novogene (Tianjin, China). Raw reads were processed with fastp [66], aligned to the Wm82.a4.v1 reference genome using Hisat2 [67], and gene and exon counts were obtained with featureCounts [68]. Exon proportion was calculated as the ratio of reads mapping to each exon relative to the total gene reads, following Li et al. [28]. TWAS were performed with the FarmCPU model [38], controlling for the top three expression PCs. Significance thresholds were adjusted using Bonferroni correction at $\alpha = 0.05$ ($p < 6.5 \times 10^{-6}$).

Genetic diversity and selective sweep analysis

Population genetic analyses were conducted using previously published SNP variation data [37]. SNPs with a missing rate > 10% or MAF < 1% were excluded. Accessions were grouped into wild soybean (*G. soja*), landraces, and improved cultivars. Nucleotide diversity (π) and fixation index (F_{st}) were calculated using VCFtools with a sliding window approach (5 kb window, 2 kb step) [69]. Tajima's *D* values were also computed to assess departures from neutral expectations and to identify genomic regions potentially subject to selective sweeps during domestication and improvement.

Plasmid construction and soybean transformation

To generate CRISPR/Cas9-mediated mutants of GmISO5, a 20-bp guide sequence targeting exon 1 was designed using the CRISPR-P online tool (https://cbi.hzau.edu.cn/crispr/) [70]. The guide sequence, driven by the U6 promoter, was inserted into a CRISPR/Cas9 expression vector under the control of the RPS5A promoter. The resulting construct was introduced into *Agrobacterium tumefaciens* strain EHA105 by electroporation and subsequently transformed into soybean cultivar Wm82 via the cotyledonary node method [71]. Transgenic plants were screened by PCR, and editing events were confirmed by Sanger sequencing of the target region. Homozygous or stable edited lines carrying frameshift mutations were advanced for phenotyping and RNA-seq analyses.

RNA-seq

Developing seeds at the R5.5 stage were harvested from wild-type (Wm82) and two independent *GmISO5* knockout lines (*iso5_L4* and *iso5_L5*), with three biological replicates per genotype. Total RNA was extracted using TRIzol reagent (Takara, Japan) following the manufacturer's protocol. mRNA was enriched using poly(A) selection and used to prepare strand-specific libraries with the NEBNext Ultra RNA Library Prep Kit (NEB, USA). Libraries were sequenced on an Illumina HiSeq 2500 platform to generate 150 bp paired-end reads (Novogene, Tianjin, China).

Adapters and low-quality reads were removed using fastp [66]. Clean reads were aligned to the soybean reference genome (Wm82.a4.v1) using HISAT2 [67], and uniquely mapped reads were summarized with featureCounts [68]. Gene expression levels were normalized as transcripts per million (TPM). Differential expression analysis was performed using DESeq2 [72], with thresholds of $|\log_2(\text{fold change})| \geq 1$ and false discovery rate (FDR) < 0.05 to define significantly differentially expressed genes (DEGs).

Yeast one-Hybrid assays (Y1H)

Promoter fragments of *GmHIDH* and *GmCHS1* were amplified from Wm82 genomic DNA and independently cloned into the pHIS2 vector to generate the pHIS2-*proGmHIDH* and pHIS2-*proGmCHS1* bait constructs. The full-length open reading frame of GmISO5 was cloned into pGADT7 to generate the prey construct. Each bait construct was co-transformed with pGADT7-GmISO5 into the yeast strain Y187 using the PEG/LiAc method. The bait constructs combined with the empty pGADT7 vector were used to assess promoter autoactivation, and pHIS2-P53/pGAD53m was used as the positive control. Transformants were initially selected on SD/-Leu/-Trp medium. To suppress bait autoactivation, transformed yeast cells were subsequently grown on SD/-Leu/-Trp/-His medium supplemented with 50 mM 3-amino-1,2,4-triazole. Yeast cultures were adjusted to equivalent densities, serially diluted, and spotted onto the selective media. Plates were incubated at 30 °C for 3–5 days [73]. Each experiment was independently repeated three times [74]. Primer sequences are listed in Supplementary Table S14.

Dual-Luciferase (LUC) reporter assay

The promoter fragments of *GmHIDH* and *GmCHS1* were independently cloned upstream of the firefly luciferase reporter gene in the pGreenII 0800-LUC vector to generate *proGmHIDH*:LUC and *proGmCHS1*:LUC reporter constructs. The Renilla luciferase gene in the same vector was used as an internal control. The full-length GmISO5 coding sequence was cloned into pBinGFP2 to generate the effector construct, and the corresponding empty vector was used as the negative control.

Reporter and effector constructs were introduced into *Agrobacterium tumefaciens* strain GV3101 and co-infiltrated into *Nicotiana benthamiana* leaves in the indicated combinations. After 24–48 h, infiltrated leaves were subjected to luminescence imaging under identical exposure conditions. Firefly and Renilla luciferase activities were quantified using a dual-luciferase reporter assay system, and promoter activity was expressed as the LUC/REN ratio. Three independent biological replicates were performed [75]. Primer sequences are listed in Supplementary Table S14.

Electrophoretic mobility shift assay (EMSA)

The full-length CDS sequence of *GmISO5* was amplified and ligated into the pET-28a vector, following a previously reported method with slight modifications [76]. The resulting recombinant construct, MBP-His-GmISO5, was transformed into competent *Escherichia coli* BL21 cells. Expression of the MBP-His-GmISO5 fusion protein was induced with 0.5 mM isopropyl β-D-1-thiogalactopyranoside (IPTG) at 28 °C for 6 h. The recombinant protein was purified using a Ni-NTA Gravity Column Protein Purification Kit (Sangon Biotech, Shanghai, China) through its His tag. The 5′-biotin-labelled DNA probes, corresponding unlabeled competitor probes and mutated probes were synthesized and annealed. EMSA was performed using the LightShift Chemiluminescent EMSA Kit (Thermo Fisher Scientific, Waltham, MA, USA) according to the manufacturer's instructions. Chemiluminescent signals were detected using a Gel Doc 2000 imaging system (Bio-Rad, Hercules, CA, USA). The primers and probe sequences used in this study are listed in Supplementary Table S14.

RNA extraction and qPCR expression analysis

For expression profiling, samples were collected from multiple tissues and developmental stages of the soybean cultivar Wm82, from Wm82 leaves before and after inoculation with soybean mosaic virus (SMV), and from developing seeds (R5.5 stage) of wild type (Wm82) and CRISPR/Cas9-generated mutants (*iso5_L4* and *iso5_L5*). Total RNA was extracted using the SteadyPure Plant RNA Extraction Kit (Accurate Biotechnology, Changsha, China) following the manufacturer's instructions. RNA integrity was verified by agarose gel electrophoresis, and purity and concentration were determined with a NanoDrop spectrophotometer. For each sample, 1 μg of total RNA was used for first-strand cDNA synthesis using the Evo M-MLV One Step RT-PCR Kit (AG11606, Accurate Biotechnology), ensuring comparable RNA input across reactions. Quantitative real-time PCR (qPCR) was performed on a CFX96 Real-Time PCR System (Bio-Rad, Hercules, USA) using SYBR Green Premix Pro Taq HS (AG11701, Accurate Biotechnology) [77]. Amplification was carried out following the manufacturer's instructions, and melting-curve analysis was used to verify product specificity. *GmActin* (*Glyma.18G290800*) served as the internal reference gene [25]. Relative expression levels were calculated using the $2^{-\Delta Ct}$ method [78]. Each experiment included three independent biological replicates, with three technical replicates per biological replicate. Gene-specific primers were designed using Primer Premier 6 [79], and their sequences are listed in Supplementary Table S14.

SMV inoculation and DAS-ELISA assay

Soybean cultivar Wm82 and CRISPR/Cas9-derived knockout lines (*iso5_L4* and *iso5_L5*) were grown in flower pots and maintained in insect-free greenhouses under controlled conditions (25 ± 2 °C, 14 h light/10 h dark photoperiod, light intensity 500–550 μmol m^{-2} s^{-1}). SMV were preserved at −80 °C and reactivated before inoculation.

Mechanical sap inoculation was performed on fully expanded unifoliolate leaves at the V1 stage, approximately 10–14 days after emergence, using carborundum-assisted rubbing with crude sap prepared from SMV-infected plants in 0.01 M phosphate-buffered saline, PBS, pH 7.4, at a 1:10 (w/v) tissue-to-buffer ratio [80].

PBS-inoculated plants served as negative controls. Phenotypic traits, including mosaic symptoms, chlorosis, plant growth, and disease severity, were monitored for at least 2 weeks post inoculation, and leaf tissues were collected at designated time points for qPCR, metabolite profiling, and DAS-ELISA assays.

DAS-ELISA was carried out using a commercial diagnostic kit (V094-R2, Nanodiagnostics, Fayetteville, United States) according to the manufacturer's protocol. Briefly, crude extracts from inoculated leaves were loaded into antibody-coated plates, incubated with alkaline phosphatase-conjugated secondary antibodies, and developed using p-nitrophenyl phosphate substrate. Absorbance at 405 nm was measured with an Infinite 200 PRO microplate reader (TECAN, Switzerland). The OD_{405} value was used as a quantitative proxy for SMV accumulation. The entire assay was conducted at 25 °C. Samples with OD_{405} values exceeding twice those of the negative controls were considered positive for SMV infection.

Statistical analysis

All statistical analyses were conducted in R (R Foundation for Statistical Computing, Vienna, Austria). Descriptive statistics were calculated for all measured traits. Differences between groups were assessed using analysis of variance (ANOVA), followed by post hoc multiple comparisons (Tukey's HSD or LSD test) where appropriate. Unless otherwise indicated, statistical significance was defined at $P < 0.05$. Exact P values, effect sizes, and 95% confidence intervals were reported where available in the figures, figure legends, and supplementary tables. For controlled molecular, expression, metabolite, and SMV experiments, data are presented as mean ± SD from at least three independent biological replicates

unless otherwise indicated. Sample sizes and statistical tests for population-level analyses are specified in the corresponding figure legends and Supplementary Table S15. Visualization of results was performed using R packages including ggplot2, ggpubr, and ComplexHeatmap, and final figure layouts were refined using Adobe Illustrator (Adobe, USA).

AI use disclosure statement

During the preparation of this work, the authors used ChatGPT (OpenAI) for language editing and grammatical refinement. The tool was used solely to improve the clarity and readability of the manuscript. The authors reviewed and edited all AI-assisted content and take full responsibility for the accuracy, integrity, and scientific validity of the published article. AI tools were not used for scientific content generation, data interpretation, reference collection, or figure preparation.

Compliance with ethics requirements

This research did not involve human participants or vertebrate animals. All plant materials used in this study were handled in accordance with institutional and national guidelines. Ethical approval was therefore not required.

Data availability

The raw RNA-seq data generated for the *GmISO5* knockout analysis have been deposited in the Genome Sequence Archive under project accession PRJCA055403. The population-level seed RNA-seq dataset used for TWAS is available under project accession PRJCA031435. The population resequencing data used for GWAS were obtained from previously published resources as cited in the Methods. All other data supporting the findings of this study are available within the article and its supplementary materials.

Author contributions

Lijuan Qiu, Xiaobo Wang, Yinghui Li designed the research; Xinkang Feng, Daqian Sun, Shiyu Guo, Hui Liu, Li Chu, Hao Zhang, Delin Li, Xueqing Wang, Tianli Ge, Shiyu Guo, Haiyang Zheng, Xiangguang Zhao, Jiajia Li, Jinyu Li performed the research and analyzed the data; Xinkang Feng, Daqian Sun drafted the manuscript; Lijuan Qiu, Yinghui Li and Xiaobo Wang edited the manuscript.

Acknowledgments

This work was supported by Inner Mongolia Autonomous Region Science and Technology Program (2023DXZD0001); National Key R&D Program of China (2021YFD1201600); the earmarked fund for CARS (CARS-04-PS01); Agricultural Science and Technology Innovation Program (ASTIP). We would like to thank Dr. Delin Li (Xianghu Laboratory) for his assistance. We also thank Dr. Hui Liu for kindly providing the SMV strains and technical support for their detection.

Declaration of competing interest

The authors declare that there are no conflicts of interest regarding the publication of this manuscript. The research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Appendix A. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jare.2026.09.011.

References

[1] Zheng H et al. GmSop20 functions as a key coordinator of the oil-to-protein ratio in soybean seeds. Adv Sci (Weinh) 2025;12:e05181.
[2] Graham PH, Vance CP. Legumes: importance and constraints to greater use. Plant Physiol 2003;131:872–7.
[3] Azam M et al. Seed isoflavone profiling of 1168 soybean accessions from major growing ecoregions in China. Food Res Int 2020;130:108957.
[4] Li Y, Zhang H. Soybean isoflavones ameliorate ischemic cardiomyopathy by activating Nrf2-mediated antioxidant responses. Food Funct 2017;8:2935–44.
[5] Laddha AP, Kulkarni YA. Pharmacokinetics, pharmacodynamics, toxicity, and formulations of daidzein: an important isoflavone. Phytother Res 2023;37:2578–604.
[6] Piovesan AC et al. Morphological and molecular effects of isoflavone and estrogens on the rat mammary gland. Bras Ginecol 2005;27:204–9.
[7] Kudou S et al. Malonyl isoflavone glycosides in soybean seeds (*Glycine max* Merrill). Agric Biol Chem 1991;55:2227–33.
[8] Zhang L et al. Genistein inhibits the stemness properties of prostate cancer cells through targeting Hedgehog-Gli1 pathway. Cancer Lett 2012;323:48–57.
[9] Lozovaya VV et al. Isoflavonoid accumulation in soybean hairy roots upon treatment with Fusarium solani. Plant Physiol Biochem 2004;42:671–9.
[10] Melnychuk M et al. Quality and safety of extracted humic substances as fundamental factor of natural regulation for the sustainable development. Agric Rural Stud 2024;2:12.
[11] Tang DYY et al. Advancing sustainable utilization of date palm, *Phoenix* dactylifera, a multifunctional resource for agro-industrial applications. Agric Rural Stud 2025;3:23.
[12] Li QY et al. Study on the effect of molecular weight on the gut microbiota fermentation properties of blackberry polysaccharides in vitro. J Agric Food Chem 2022;70:11245–57.
[13] Chen X, Chen C, Fu X. Hypoglycemic effect of the polysaccharides from Astragalus membranaceus on type 2 diabetic mice based on the "gut microbiota-mucosal barrier". Food Funct 2022;13:10121–33.
[14] Ge C et al. Cyanoglycosides isolated from Moringa oleifera seeds inhibited PFKFB3/TGF-β1/smads pathway to alleviate diabetic nephropathy through driving metabolic reprogramming. Pharm Biol 2026;64:130–42.
[15] Yu O et al. Metabolic engineering to increase isoflavone biosynthesis in soybean seed. Phytochemistry 2003;63:753–63.
[16] Jung W et al. Identification and expression of isoflavone synthase, the key enzyme for biosynthesis of isoflavones in legumes. Nat Biotechnol 2000;18:208–12.
[17] Wu D et al. Identification of a candidate gene associated with isoflavone content in soybean seeds using genome-wide association and linkage mapping. Plant J 2020;104:950–63.
[18] Anguraj Vadivel AK, McDowell T, Renaud JB, Dhaubhadel S. A combinatorial action of GmMYB176 and GmbZIP5 controls isoflavonoid biosynthesis in soybean (*Glycine max*). Commun Biol 2021;4:356.
[19] Bian S et al. Soybean CCA1-like MYB transcription factor GmMYB133 modulates isoflavonoid biosynthesis. Biochem Biophys Res Commun 2018;507:324–9.
[20] Chu S et al. An R2R3-type MYB transcription factor, GmMYB29, regulates isoflavone biosynthesis in soybean. PLoS Genet 2017;13:e1006770.
[21] Jiang H et al. Multi-omics analysis identified the GmUGT88A1 gene, which coordinately regulates soybean resistance to cyst nematode and isoflavone content. Plant Biotechnol J 2025;23:1291–307.
[22] Feng Y et al. Dual-function C2H2-type zinc-finger transcription factor GmZFP7 contributes to isoflavone accumulation in soybean. New Phytol 2023;237:1794–809.
[23] Liu Y et al. An R2R3-type MYB transcription factor, GmMYB77, negatively regulates isoflavone accumulation in soybean [*Glycine max* (L.) Merr.]. Plant Biotechnol J 2025;23:824–38.
[24] Yan J et al. The soybean R2R3 MYB transcription factor GmMYB100 negatively regulates plant flavonoid biosynthesis. Plant Mol Biol 2015;89:35–48.
[25] Song Z et al. The GmSTF1/2-GmBBX4 negative feedback loop acts downstream of blue-light photoreceptors to regulate isoflavonoid biosynthesis in soybean. Plant Commun 2024;5:100730.
[26] Qin C et al. PH13 improves soybean shade traits and enhances yield for high-density planting at high latitudes. Nat Commun 2023;14:6813.
[27] Zheng H et al. GmSop20 Functions as a Key Coordinator of the Oil-To-Protein Ratio in soybean Seeds. Adv Sci (Weinh) 2025;e05181.
[28] Li D et al. TWAS facilitates gene-scale trait genetic dissection through gene expression, structural variations, and alternative splicing in soybean. Plant Commun 2024;5:101010.
[29] Wigge PA et al. Integration of spatial and temporal information during floral induction in Arabidopsis. Science 2005;309:1056–9.
[30] Xi W, Liu C, Hou X, Yu H. MOTHER OF FT AND TFL1 regulates seed germination through a negative feedback loop modulating ABA signaling in Arabidopsis. Plant Cell 2010;22:1733–48.

[31] Taoka K et al. 14-3-3 proteins act as intracellular receptors for rice Hd3a florigen. Nature 2011;476:332–5.
[32] Vaistij FE et al. MOTHER-OF-FT-AND-TFL1 represses seed germination under far-red light by modulating phytohormone responses in *Arabidopsis thaliana*. PNAS 2018;115:8442–7.
[33] Zhang B, Li C, Li Y, Yu H. Mobile TERMINAL FLOWER1 determines seed size in Arabidopsis. Nat Plants 2020;6:1146–57.
[34] Chen Y et al. Nuclear translocation of OsMFT1 that is impeded by OsFTIP1 promotes drought tolerance in rice. Mol Plant 2021;14:1297–311.
[35] Cai Z et al. MOTHER-OF-FT-AND-TFL1 regulates the seed oil and protein content in soybean. New Phytol 2023;239:905–19.
[36] Duan Z et al. Natural allelic variation of GmST05 controlling seed size and quality in soybean. Plant Biotechnol J 2022;20:1807–18.
[37] Li YH et al. Genome-wide signatures of the geographic expansion and breeding of soybean. Sci China Life Sci 2023;66:350–65.
[38] Liu X, Huang M, Fan B, Buckler ES, Zhang Z. Iterative usage of fixed and random effect models for powerful and efficient genome-wide association studies. PLoS Genet 2016;12:e1005767.
[39] Yoshikawa T et al. Transgressive segregation of isoflavone contents under the control of four QTLs in a cross between distantly related soybean varieties. Breed Sci 2010;60:243–54.
[40] Primomo VS et al. Mapping QTL for individual and total isoflavone content in soybean seeds. Crop Sci 2005;45:2454–64.
[41] Yang K et al. Novel major quantitative trait loci regulating the content of isoflavone in soybean seeds. Genes Genomics 2011;33:685–92.
[42] Gutierrez-Gonzalez JJ et al. Intricate environment-modulated genetic networks control isoflavone accumulation in soybean seeds. BMC Plant Biol 2010;10:105.
[43] Imaizumi R et al. Crystal structure of chalcone synthase, a key enzyme for isoflavonoid biosynthesis in soybean. Proteins 2020.
[44] Dao TT, Linthorst HJ, Verpoorte R. Chalcone synthase and its functions in plant resistance. Phytochem Rev 2011;10:397–412.
[45] Mameda R, Waki T, Kawai Y, Takahashi S, Nakayama T. Involvement of chalcone reductase in the soybean isoflavone metabolon: identification of GmCHR5, which interacts with 2-hydroxyisoflavanone synthase. Plant J 2018;96:56–74.
[46] Hamberger B, Hahlbrock K. The 4-coumarate:CoA ligase gene family in *Arabidopsis thaliana* comprises one rare, sinapate-activating and three commonly occurring isoenzymes. Proc Natl Acad Sci USA 2004;101:2209–14.
[47] Nesi N, Jond C, Debeaujon I, Caboche M, Lepiniec L. The Arabidopsis TT2 gene encodes an R2R3 MYB domain protein that acts as a key determinant for proanthocyanidin accumulation in developing seed. Plant Cell 2001;13:2099–114.
[48] Zhao G et al. Molecular loci associated with seed isoflavone content may underlie resistance to soybean pod borer (Leguminivora glycinivorella). Plant Breed 2015;134:78–84.
[49] Meng F, Han Y, Teng W, Li Y, Li W. QTL underlying the resistance to soybean aphid (Aphis glycines Matsumura) through isoflavone-mediated antibiosis in soybean cultivar 'Zhongdou 27'. Theor Appl Genet 2011;123:1459–65.
[50] Tian M et al. Metabolomics navigates natural variation in pathogen-induced secondary metabolism across soybean cultivar populations. Proc Natl Acad Sci USA 2025;122:e2505532122.
[51] Zhang P et al. Multiplex CRISPR/Cas9-mediated metabolic engineering increases soya bean isoflavone content and resistance to soya bean mosaic virus. Plant Biotechnol J 2020;18:1384–95.
[52] Ye Y et al. A new isoflavonoid from the stems of *Nicotiana tabacum* and its anti-tobacco mosaic virus activity. Asian J Chem 2012;24:5393–4.
[53] Liu X et al. Over-expression of GmMYB39 leads to an inhibition of the isoflavonoid biosynthesis in soybean (*Glycine max*. L). Plant Biotechnol Rep 2013;7:445–55.
[54] Han X et al. GmMYB58 and GmMYB205 are seed-specific activators for isoflavonoid biosynthesis in *Glycine max*. Plant Cell Rep 2017;36:1889–902.
[55] Polturak G et al. Discovery of isoflavone phytoalexins in wheat reveals an alternative route to isoflavonoid biosynthesis. Nat Commun 2023;14:6977.
[56] Picmanov M, Honys D, Koblovsk R, Lapcik O. Isoflavone synthase genes in legumes and non-leguminous plants: isoflavone synthase. In: 2012 International Conference on Biomedical Engineering and Biotechnology. p. 344–7.
[57] Yu O, McGonigle B. Metabolic engineering of isoflavone biosynthesis. In: Advances in Agronomy. Academic Press; 2005. p. 147–90.
[58] Han L et al. A multi-omics integrative network map of maize. Nat Genet 2023;55:144–53.
[59] Tang S et al. Genome- and transcriptome-wide association studies provide insights into the genetic basis of natural variation of seed oil content in *Brassica napus*. Mol Plant 2021;14:470–87.
[60] Tan Z et al. Genome- and transcriptome-wide association studies reveal the genetic basis and the breeding history of seed glucosinolate content in *Brassica napus*. Plant Biotechnol J 2022;20:211–25.
[61] Zhang Y et al. Multi-omics analysis dissects the genetic architecture of seed coat content in *Brassica napus*. Genome Biol 2022;23:86.
[62] Yuan X et al. Integrative omics analysis elucidates the genetic basis underlying seed weight and oil content in soybean. Plant Cell 2024;36:2160–75.
[63] Zhao T et al. Population-wide DNA methylation polymorphisms at single-nucleotide resolution in 207 cotton accessions reveal epigenomic contributions to complex traits. Cell Res 2024;34:859–72.
[64] Sun J-M et al. Rapid HPLC method for determination of 12 isoflavone components in soybean seeds. Agric Sci China 2011;10:70–7.
[65] Chang CC et al. Second-generation PLINK: rising to the challenge of larger and richer datasets. GigaScience 2015;4:7.
[66] Chen S, Zhou Y, Chen Y, Gu J. fastp: an ultra-fast all-in-one FASTQ preprocessor. Bioinformatics 2018;34:i884–90.
[67] Kim D, Langmead B, Salzberg SL. HISAT: a fast spliced aligner with low memory requirements. Nat Methods 2015;12:357–60.
[68] Liao Y, Smyth GK, Shi W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics 2014;30:923–30.
[69] Danecek P et al. The variant call format and VCFtools. Bioinformatics 2011;27:2156–8.
[70] Lei Y et al. CRISPR-P: a web tool for synthetic single-guide RNA design of CRISPR-system in plants. Mol Plant 2014;7:1494–6.
[71] Li S et al. Optimization of agrobacterium-mediated transformation in soybean. Front Plant Sci 2017;8–2017.
[72] Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol 2014;15:550.
[73] Liu T et al. PgMYB2, a MeJA-responsive transcription factor, positively regulates the dammarenediol synthase gene expression in *Panax ginseng*. Int J Mol Sci 2019;20.
[74] Lv X et al. Integrated metabolomics and transcriptomics analyses highlight the flavonoid compounds response to alkaline salt stress in *Glycyrrhiza uralensis* leaves. J Agric Food Chem 2024;72:5477–90.
[75] Li X et al. Integrative physiology and transcriptome reveal salt-tolerance differences between two licorice species: Ion transport, Casparian strip formation and flavonoids biosynthesis. BMC Plant Biol 2024;24:272.
[76] Wang C et al. Metabolomics and transcriptomic analysis revealed the response mechanism of maize to saline-alkali stress. Plant Biotechnol J 2025.
[77] Abubakar AS et al. Comprehensive analysis of WUSCEL-related homeobox gene family in Ramie (*Boehmeria nivea*) indicates its potential role in adventitious root development. Biology (Basel) 2023;12.
[78] Livak KJ, Schmittgen TD. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-ΔΔC(T)) Method. Methods 2001;25:402–8.
[79] Singh VK, Mangalam AK, Dwivedi S, Naik S. Primer premier: program for design of degenerate primers from a protein sequence. Biotechniques 1998;24:318–9.
[80] Yin J et al. A cell wall-localized NLR confers resistance to soybean mosaic virus by recognizing viral-encoded cylindrical inclusion protein. Mol Plant 2021;14:1881–900.